

Commonalities between perception and cognition

Michela C. Tacca*

Department of Philosophy, Heinrich-Heine University Düsseldorf, Düsseldorf, Germany

Edited by:

Arnon Cahen, Ben-Gurion University of the Negev, Israel

Reviewed by:

Arnon Cahen, Ben-Gurion University of the Negev, Israel
Ellen Fridland, Humboldt University of Berlin, Germany

***Correspondence:**

Michela C. Tacca, Department of Philosophy, Heinrich-Heine University, Universitätsstr. 1, 40225 Düsseldorf, Germany.
e-mail: michela.tacca@gmail.com

Perception and cognition are highly interrelated. Given the influence that these systems exert on one another, it is important to explain how perceptual representations and cognitive representations interact. In this paper, I analyze the similarities between visual perceptual representations and cognitive representations in terms of their structural properties and content. Specifically, I argue that the spatial structure underlying visual object representation displays systematicity – a property that is considered to be characteristic of propositional cognitive representations. To this end, I propose a logical characterization of visual feature binding as described by Treisman's Feature Integration Theory and argue that systematicity is not only a property of language-like representations, but also of spatially organized visual representations. Furthermore, I argue that if systematicity is taken to be a criterion to distinguish between conceptual and non-conceptual representations, then visual representations, that display systematicity, might count as an early type of conceptual representations. Showing these analogies between visual perception and cognition is an important step toward understanding the interface between the two systems. The ideas here presented might also set the stage for new empirical studies that directly compare binding (and other relational operations) in visual perception and higher cognition.

Keywords: systematicity, generality constraint, conceptual content, non-conceptual content, attention

INTRODUCTION

Perception and cognition are tightly related. Perceptual information guides our decisions and actions, and shapes our beliefs. At the same time our knowledge influences the way we perceive the world (Brewer and Lambert, 2001). To the extent that perception and cognition seem to share information, it seems there is no sharp division between the realm of cognitive abilities and that of perceptual abilities. An example is visual perception. Visual processing is composed of different stages (Marr, 1982): early, intermediate, and late vision. Roughly, at early stages of the visual system, processes like segregation of figure from background, border detection, and the detection of basic features (e.g., color, orientation, motion components) occur. This information reaches intermediate stages, where it is combined into a temporary representation of an object. At later stages, the temporary object representation is matched with previous object shapes stored in long-term visual memory to achieve visual object identification and recognition. While early visual processes are largely automatic and independent of cognitive factors, late visual stages are more influenced by our knowledge (Raftopoulos, this issue). Examples of cognitive influence on how we perceive the world – that modulates late vision – are visual search and attention (Treisman, 1993). Knowing the color or shape of an object helps a person to quickly identify that particular object in a cluttered visual scene (Wolfe and Horowitz, 2004). Phenomena like visual search highlight the fact that visual perception at later stages depends on both sensory and cognitive factors. Late vision is at what philosophers call the personal level: we have conscious access to information represented at this stage and we can exploit it for action planning and thinking (Lamme, 2003; Block, 2005). This is apparently not the case for early visual stages, which

occur at a subpersonal level, without a person being aware of the information being processed at that stage. Intermediate stages, on the other hand, are probably accessible at a personal level. The degree of representational awareness occurring at this stage is commonly identified with phenomenal consciousness (Lamme, 2003; Raftopoulos and Mueller, 2006): we get a gist of the perceived scene, but it is not possible to retrieve detailed information of the objects' features. It is a matter of debate to what extent intermediate stages of visual processing are influenced by our knowledge (i.e., are cognitively penetrable). Some authors argue that those stages are purely visual (Raftopoulos and Mueller, 2006) and that the transition between pure perception to cognition occurs only at later visual stages, when temporary object representations are matched for recognition and identification. In this paper, I will not propose an argument for whether early and intermediate stages of visual perception are cognitively penetrable. However, I would like to stress that some of the common properties between visual perception and cognition that I will consider already occur at intermediate stages, thus, casting doubt on the claim that mid-level vision is purely perceptual.

Cognitive information influences perceptual processes, but, at the same time, cognitive processes depend on perceptual information (Goldstone and Barsalou, 1998). Recent work in philosophy brought new vigor to the hypothesis originally proposed by British Empiricists that cognition is inherently perceptual (Prinz, 2002): cognitive/conceptual tasks have their roots in perception and they rely on perceptual mechanisms for their processing. Such theoretical proposals are supported by empirical findings from psychology. Work on concept acquisition shows that functions (e.g., categorization, inference) that are associated with cognition have their

basis in perceptual systems (Barsalou, 1999) and that perceptual and conceptual processes share common mechanisms (Pecher et al., 2004). The basic hypothesis is that a concept is represented by means of a simulation at the sensory level of an experience of that to which the concept truly applies. For example, to represent the concept APPLE¹, perceptual systems for vision, action, and touch partially produce the experience of a particular apple. Taken together, work on the influence of knowledge on the character of one's perceptual experience and on perceptual information shaping one's conceptual abilities provides evidence for perception and cognition being related systems.

Though it seems to be common ground that cognitive and perceptual representations influence each other, they are not taken to be the same kind of representations. Neurophysiological studies distinguish different functional areas for sensory and cognitive systems. Those areas process specific inputs and specialize in different kinds of information processes (Zeki, 1978; Felleman and Van Essen, 1991). And distinct sensory areas can be treated as separate modules (Barrett, 2005) that deal with their specific representational primitives.

From a philosophical point of view, visual perception and cognition process information by means of representations that differ in both their structure and content (Heck, 2007; Fodor, 2008). One of the main characteristics of cognitive states, paradigmatically of thoughts, is that they have a propositional combinatorial structure that satisfies the requirement of the Generality Constraint (Evans, 1982). The Generality Constraint describes the pervasive ability of humans to entertain certain thoughts that they have never had before on the basis of having entertained the components of these new thoughts in other preceding situations. For example, from the fact that a person can think that the sky is blue and the car is gray, she can also think that the sky is gray and the car is blue, even if she has never had this thought before. The new thought depends on her conceptual ability to combine already acquired concepts in different ways. This regularity of human thinking is explained by appealing to the fact that thoughts are mental representations with a sentential combinatorial structure (Fodor, 1975). Thoughts are built up by combining primitive constituents according to propositional rules. The thought 'the car is gray' depends on the tokening and combination of the concepts CAR and GRAY and the rule of composition for the verb 'to be.' Recombination of concepts in cognitive processes displays a constituent structure. The constituent structure of thought is such that whenever a complex representation is tokened its constituents are simultaneously tokened. Failure to represent car or grayness leads to failure to represent that the car is gray. The appeal to the constituent structure of cognitive representations allows us to explain a further property of these representations: their systematicity (Fodor and Pylyshyn, 1988). Systematicity, similar to the Generality Constraint, describes the human ability to entertain semantically related thoughts. For example, the ability to entertain a certain thought about cars is connected to the ability to entertain certain other thoughts about cars: thoughts like 'the car is

gray' and 'the car is blue' share the same constituent 'car.' That is, the semantic systematicity of thought is explained by postulating a system of representations with a combinatorial syntax.

Systematic recombinations are necessary to satisfy the Generality Constraint but not sufficient. In fact, systematicity is a weaker requirement than the Generality Constraint since it lacks the "generality" part. According to the Generality Constraint, once a thinker can entertain a thought, elements of this thought could be in principle indefinitely recombined with every other appropriate concept that a person possesses. This requirement is not part of systematicity, since it leaves open whether it is in principle possible that a finite type of systematicity exists (Fodor and Pylyshyn, 1988). For what concerns the analysis of the structure of visual representations, I will mostly focus on whether those representations implement a systematic structure of constituents. I will then discuss the "generality" requirement in the analysis of the content of visual representations.

Acceptance of the Generality Constraint, or the weaker systematicity requirement, also affects how we characterize the content of cognitive and perceptual representations. Philosophers distinguish between two types of content: conceptual and non-conceptual content (Evans, 1982; Bermúdez, 2007). Typical cases of mental states with conceptual content are cognitive mental states, like thought, belief, desire, and so on: their content – what is thought, believed, and desired – is a function of the concepts a person possesses, where concepts are taken to be the constituents of thoughts and other cognitive states. Mental states with non-conceptual content, on the other hand, are states the having of which does not depend on the subject's possessing any of the concepts required to specify the content of that state. Perception, both personal and subpersonal, is considered a paradigmatic example of states with non-conceptual content. In other words, to have the thought that an apple is red, one has to possess the concepts involved in that thought, but to have a perceptual experience characteristic of seeing a red apple one does not need to possess the concepts involved in the specification.

One way of distinguishing conceptual and non-conceptual content appeals to a mental representation's satisfaction of the systematicity requirement (Toribio, 2008; Camp, 2009). It has been argued that perceptual representations, specifically visual representations, do not satisfy the requirement of systematicity, and, hence, unlike cognitive representations, do not have conceptual content (Heck, 2007). The argument is based on the idea that visual representations have a pictorial nature. Pictorial theories equate visual representations to images or maps. Like images or maps, visual representations are spatially characterized: at each point in an image or map a specific trait (color, shape, etc.) occurs. Furthermore, like images or maps, visual representations have a holistic character. Unlike cognitive representations, there is no unique structured propositional representation that determines the content of a visual representation. There are many distinct possible decompositions of the same image, such that it is impossible to both identify which are its constituent parts and disentangle the role of these parts in the building up of the pictorial representation. Thus, visual representations, like maps, seemingly lack the syntactic structure of constituents typical of cognitive representations. The lack of a constituent structure entails that visual

¹I follow the common practice in philosophy to capitalize terms that refer to concepts.

representations are not systematic. Satisfying systematicity is a necessary condition on satisfying the Generality Constraint. For the reasons above, visual representations do not seem to satisfy systematicity, and hence the Generality Constraint. Therefore, they have a content of a different kind than the content of cognitive representations: they have non-conceptual content.

If visual perception and cognition do indeed have different structural properties and content, then it becomes difficult to understand how perceptual representations are “translated” into cognitive representations. This is both an empirical and theoretical question. From the philosophical point of view, finding out the relationship between perception and cognition will be of benefit to explain phenomena as different as concept formation and acquisition, belief justification, and demonstrative thinking, each of which partly depends on perceptual information.

In this paper, I will focus on commonalities between visual perception and cognition that might help explain the communication between those systems. In the first part, I will show that the spatial recombination underlying visual object recognition satisfies the requirement of systematicity. The analysis will take into account the so-called Feature Integration Theory (Treisman and Gelade, 1980); a model that explains visual object representation by considering the spatial nature of visual representations. Although Feature Integration Theory characterizes visual representations as spatially organized, it differs from pictorial theories of visual representations, since it does not commit to the view that visual representations are holistic. In fact, visual representations can be seen as states of the visual system that can be neurally specified, such that each part of an object representation can be spelled out by considering the different neuronal activations (Treisman and Gelade, 1980; Goldstone and Barsalou, 1998). Each neuronal activation roughly corresponds to a part, or primitive constituent, of the representation. Thus, one can decompose an object representation into its primitive constituents and analyze whether a systematic structure of constituents is displayed by visual spatial recombinations (Tacca, 2010). In the second part, I will argue against the claim that visual representations have non-conceptual content. Based on the analysis in the first part of the paper, I will propose that, if one takes systematicity to be a necessary requirement for having conceptual content, visual representations might be an *early* type of conceptual representations. I conclude that understanding the link between perception and cognition requires considering whether they satisfy common requirements in terms of structure and content. These similarities might be at the basis of the translation of perceptual representations into cognitive representation and elucidate the mechanism of their interaction.

PRIMITIVE VISUAL FEATURES AND THE BINDING PROBLEM

Recombination in cognitive processes depends on operations on primitive constituents. A primitive constituent is an entity that corresponds to the smallest meaningful representation carrying relevant information for the processing of more complex representations. Different theories posit different types of primitive constituents (Smolensky, 1990; Fodor, 1998). However, there is agreement that the primitive mental representations involved in thought and other cognitive processes, like belief and desire, are concepts. According to an atomistic perspective, concepts cannot

be further decomposed into more primitive elements and as such they are the building blocks of thoughts (Fodor, 1975). However, others have argued that concepts can be further decomposed into their perceptual components (e.g., Barsalou, 1999). For example, the concept APPLE can be decomposed into its constituent concepts: COLOR, TEXTURE, SHAPE, etc. At the same time, each part can be further decomposed into more elementary constituents like GREEN, BROWN, SMOOTH, and ROUND. Those elementary constituents are taken to be symbolic perceptual representations stored at late perceptual stages that become part of cognitive recombinations. Therefore, they share with cognitive representations systematicity, compositionality, and productivity (Barsalou, 1999). In the following, I will show that intermediate visual representations that contribute to object perception but are not yet stored at late visual stages also display systematicity.

The hypothesis that concepts have a structure of constituents that involves perceptual representations is based on anatomical, physiological, and psychophysical evidence for the existence of distinct representations for primitive visual features. Neurobiological (Zeki, 1978; Livingstone and Hubel, 1988; Felleman and Van Essen, 1991) and psychophysical studies (Treisman and Gelade, 1980) report the existence in visual areas of so-called feature maps. Feature maps code for specific object features, like color, motion, and orientation. They are also topographically organized; namely, they represent a specific feature and the specific location in which the feature occurs in the visual field. Thus, any visual object we perceive is first decomposed into its primitive components and only later those components are recombined into a coherent object representation. But what makes color, motion, and orientation count as primitive features not further decomposable? To provide an answer to this question is important, since if we can show that there is an empirically reasonable standard for primitive recombinable features, then we can challenge one of the central motivation for thinking that visual perception does not display systematicity and that the content of visual representations is non-conceptual; namely, the claim of pictorial theories for which there is no unique decompositions of visual representations into a proper structure of constituents.

The definition of a primitive visual feature not further decomposable depends on experimental consideration (Wolfe, 1998). First, a primitive feature allows for efficient visual search when embedded in a cluttered scene of unlike distracters. The efficiency of visual search is indicated by the so-called “pop-out” of the target that is independent of how many items are present in the visual field. Second, a primitive feature supports effortless texture segregation. For example, a region of vertical lines in a field of horizontal lines will be immediately segregated from the background and perceived as a figure. Color, orientation, and motion justify the criteria of efficient search and effortless segmentation, and are, thus, primitive features. Furthermore, these features are represented by different visual cortical areas, each of which is retinotopically organized. Taken together, neurophysiological and psychophysical findings uncover the fact that visual features are the primitive constituents of visual object representations.

Once primitive visual features have been individuated, the subsequent main question is how those features are combined. In light of the complexity of natural visual scenes, it is striking that features

are almost never miscombined in our perception. In fact, this is even remarkable for the simplest possible scenes, such as one with a red-horizontal bar and a green-vertical bar and another one with a green-horizontal bar and a red-vertical bar. These scenes contain identical features that are combined in different ways. The challenge consists in individuating objects by their unique combination of features, so as to distinguish, for example, the red-vertical bar from the green-horizontal bar. Jackson (1977) described the problem of feature recombination as the Many-Property problem. Research in vision science has approached this problem under the label of “binding problem” (Roskies, 1999). An example of what the binding problem involves comes from studies of visual conjunction search (Treisman and Gelade, 1980). A typical case of feature integration is to show a subject a scene in which red-vertical bars, red-horizontal bars, green-horizontal bars, and one green-vertical bar are presented together. The subject is asked to identify the green-vertical bar. In order to detect the right target, something like a comparison between the right orientation and the right color has to occur. It has been shown that in the case of identification of objects that share different features (orientation and color in the example case) selective attention is at play (Treisman, 1996). Further evidence for the binding problem being solved by an attentional mechanism comes from studies of illusory conjunctions in healthy subjects (Treisman and Schmidt, 1982) and patients suffering from Balint’s syndrome (Robertson, 2003). Healthy subjects are asked in a laboratory setting to report properties of visually presented stimuli under high attentional load. Results show that they report a high number of illusory conjunctions. For example, when shown a screen with blue squares and red triangles, they report wrong recombinations of presented features, e.g., a blue triangle. A high rate of illusory conjunctions occurs if similar experiments are performed with Balint’s syndrome patients. These patients suffer, among other things, from an attentional disruption, providing more evidence for the role of attention in successful binding.

The reported findings support the so-called Feature Integration Theory (Treisman and Gelade, 1980). Feature Integration Theory is one of the most influential models of visual feature binding that considers the role of attention and the spatial layout of feature maps as the basic ingredients for successful feature binding. Other influential models have been proposed for explaining the binding process, such as the hypothesis of binding by synchrony that considers synchronized neuronal mechanism as the basic binding mechanism (Engel et al., 1991). Furthermore, besides the spatial-attentional mechanism posited by Feature Integration Theory, also object-based attention might be necessary to integrate features (Blaser et al., 2000). The hypotheses of binding by spatial attention, synchrony, and of the role of object-based attention are not mutually exclusive (Tacca, 2010). It might be that all these factors are at play during the binding process. Indeed, empirical studies show the relation between spatial attention and synchrony (Fries et al., 2001) and between object-based and spatial attention (Scholl, 2009) in building up an object representation. Here, I will only focus on the role of spatial attention to bind features, in order to show that spatial representations display systematicity in a way similar to cognitive-sentential representations.

According to Feature Integration Theory, selective attention acts as the active binding mechanism. Whenever a person focuses

her attention on a specific object location in the visual field, the features at that location are represented in the corresponding location in the feature maps. By selecting all the features occupying a specific location, attention integrates these into a coherent object representation. More specifically, the focus of attention selects an object location within a topographically organized master map of location (Treisman, 1993) or saliency-map (Koch and Ullman, 1985). This saliency-map represents the saliency of objects at each location of the visual field, because it combines the information about all features’ saliency from all the specific feature maps, which it receives via topographically organized connections from the feature maps. Within each feature map, the saliency at a given location is determined by two classes of factors (Wolfe, 1998): (i) bottom-up saliency, that is, the local feature gradient (Koch and Ullman, 1985); and (ii) top-down factors, like the match between a stimulus feature and the features of the object that a person is currently searching for (Wolfe, 1998).

Independently of whether the saliency of individual locations is governed by bottom-up or top-down factors, the saliency representation in the saliency-map is always generated by combining the outputs from feature maps in a fashion that preserves topography. That is, the saliency-map receives information about the different object locations – suppose that locations are indexed with i , l , m , n , etc. – and their conspicuity values from distinct feature maps. If location _{i} , signaled by the feature map _{α} , is the same as location _{l} ($i = l$), signaled by the feature map _{β} , they will activate the same portion of the saliency-map. The saliency value of this location will then depend on the conspicuity of both location _{i} and location _{l} . The saliency-map only codes for saliency at a given location. Thus, the saliency-map represents the locations of objects but has no information about which features occur at those locations. In order to recover which features determine the object’s shape and surface, information within the topographic feature maps has to be selected for binding and further processing of object identity. A “winner take-all” mechanism selects the location in the saliency-map that is the most salient at any given moment (Koch and Ullman, 1985). This determines where the focus of attention will next move. Via topographically ordered feedback connections from the saliency maps to the corresponding locations in the feature maps, the features at that location (e.g., features occurring at both location _{i} in the feature map _{α} and location _{l} in the feature map _{β} , since $i = l$) are jointly selected for further processing, and, in this way, bound. These integrated features are stored as temporary representations – called by some authors an object-file (Kahneman et al., 1992) – in which their constituting information of location is indexed. Hence, in models based on Feature Integration Theory, the representation of objects’ locations is fundamental for integrating their features.

In this framework, the difference between saliency being governed by bottom-up or top-down factors amounts to the distinction between exogenous and endogenous attention. Exogenous attention is governed by stimulus property: it is attracted by the conspicuity of an object in the perceived scene. If you are attending a seminar and a fly suddenly enters into the room, you will immediately spot and follow it. No matter how much you are interested in the seminar. Endogenous attention is governed by a subject’s tasks and plans. You want to wear your favorite

pullover and you go through the content of your messy closet to find it. You will drive your attention to the location where you thought the pullover should be, if you are lucky your search is over, but, as often happens, you will have to scan through different locations before you can find it among all the other similar cloths.

Note that, in the sequence of processes postulated by Feature Integration Theory, the binding process is separate from the representation of location saliency. In principle, binding can be disrupted without a disruption of the saliency representation. Thus, in this framework, attention and binding can come apart. To illustrate a scenario in which such dissociation occurs, let us assume that we selectively interrupt the feedback connections from the saliency-map to the feature maps, leaving everything else intact. Then, there will still be a most salient location selected in the saliency-map and only the final process in the above sequence will be disrupted. Suppose that the perceived scene is one with a green-vertical bar and a red-horizontal bar. Object features are represented in feature maps according to their location: green_i, vertical_j, red_m, and horizontal_n. Information about feature locations is sent to the saliency-map, which computes the most salient location. In the saliency-map, location_i and location_j activate the same area (location_i = location_j), since they bring information about the same object location, and location_m and location_n activate the same area (location_m = location_n) that is different from the location of the object signaled by location_i and location_j. Suppose that the location of *i* and *j* is the most salient, then attention will be directed to this location and a signal to select features “indexed” *i* and *j* will be sent to the feature maps. Since the feedback connections from the saliency-map to the feature maps are disrupted, features in the feature maps belonging to the same location cannot be selected. The feature maps will encode for features and their locations, but there is no selective feedback signal that routes only those features from the selected location to the next step of object processing that binds them. This might result in perceptual misbinding because features from many locations are spuriously sent on to higher-level object processing. In fact, one possibility is that psychophysical manipulations leading to illusory conjunctions (Treisman and Gelade, 1980) work by interrupting the feedback from the saliency-map to the feature maps, just as in this thought experiment. For proper binding, information about features occupying the same identical location has to be routed from the feature maps to higher processing stages.

Also note that, even with disturbed saliency representation, and thus disturbed attention, some feature binding (even if erroneous) occurs. An empirical example for this can be found in Balint's syndrome patients. Spatial attention in these patients is disrupted, yet they still report a (wrong) recombination of features. Thus, even without spatial attention, some erroneous form of binding can occur. The fact that attentional selection and feature binding are tightly related, yet distinct processes, is of importance for the analysis of the binding process in the logical terms that are proposed in the next section.

Briefly, the main ingredients of Feature Integration Theory are the representation of primitive features, their spatial location, and attention. The interaction of these elements gives rise to the

perception of objects in a scene in which features are correctly conjoint. This might solve the Many-Property or binding problem at least in the case of visual object representation.

SYSTEMATIC RECOMBINATION OF SPATIALLY ORGANIZED REPRESENTATIONS

Models based on Feature Integration Theory describe visual object representations as the outcomes of recombinations of primitive visual constituents. This contrasts with pictorial theories of visual perception in philosophy (e.g., Heck, 2007; Fodor, 2008) that argue that visual representations have a holistic nature. Visual representations, like images or maps, can be decomposed in many different ways: to each visual representation might correspond a different decomposition of constituents. That means that any kind of decomposition of a visual representation into its constituents makes the same contribution to the final object representation. The decomposition of, for example, a visual representation of a flower into (petals, stem, leaves) is as good as the decomposition (part of petal₁, roots, sepal, stalk). Therefore, visual representations are, unlike cognitive representations, not canonically decomposable (Fodor, 2007): while the decomposition of a thought representation allows only a unique decomposition – e.g., ‘John loves Mary’ decomposes into JOHN, LOVES, and MARY – iconic representations have infinitely many decompositions, none of which is canonical. Having a structure of primitive constituents depends on the individuation of the unique parts of a canonical decomposition. Since visual representations seemingly fail to canonically decompose, they lack a structure of primitive constituents. To implement a structure of constituents is a pre-requisite for explaining the systematic behavior of cognitive processes. The relation of constituency is defined as a mereological relation; namely, as a relation of parts to whole (Fodor and McLaughlin, 1990): every time the expression *E* is tokened, its constituents $\langle e_1, \dots, e_n \rangle$ are tokened, too. In a classical account of thought processes, systematicity results from processes that are sensitive to the structure of constituents: the ability to entertain related thoughts depends on the fact that different combinations of constituents have the same syntactic structure. As an example, the thoughts ‘John loves Mary’ and ‘Mary loves John’ share the same structure, even if the constituents are differently arranged. According to the pictorialists, because of the holistic character of visual representations, those representations fail to implement such a structure of constituents, and, as a consequence, they do not display systematicity.

Empirical evidence casts doubt on the main assumption of pictorial theories: that perceptual representations have a holistic character, and therefore lack systematicity. Evidence from vision science shows that visual object representations depend on the recombination of neuronally specified primitive features. These features can be uniquely determined in terms of neuronal activations, and they are represented in distinct feature maps. Experimental considerations make clear that features represented in the feature maps are primitive and not further decomposable. Object representations then depend on the spatial recombination of those features. It seems plausible that such recombinations display systematicity; namely that visual scenes that are structurally related (e.g., to see a red-vertical bar to the left of a green-horizontal

bar and *vice versa*) share the same primitive visual features (i.e., 'green,' 'horizontal,' 'red,' and 'vertical'). In order to show that this is indeed the case, one has to first argue that visual representations implement a mereological structure of constituents, such that every time an object representation is tokened its primitive features are tokened, too; and, second, that the visual system implements a systematic structure of constituents; namely, that visual features make the same contribution in structurally related visual scenes.

The analysis of the type of structure implemented in the process of binding by attention, as described by Feature Integration Theory, can be given in logical terms (Clark, 2004a; Tacca, 2010). Binding involves predication and identity: features are considered to be the predicates of the same sensory individual that, in the case of Feature Integration Theory, is the object location. The reason for introducing identity is that a pure conjunction of terms might lead to different representations of the same scene, each of which would be valid. Consider, for example, the simple visual scene with a red-vertical bar and a green-horizontal bar. Its decomposition only by means of conjunction would be: (red and vertical and green and horizontal). The recombination of those features could lead to two distinct visual scenes: one in which there are a red-vertical bar and a green-horizontal bar, and one in which there are a red-horizontal bar and a green-vertical bar. This kind of ambiguity does not occur in object perception. The binding process normally produces a unique representation of the objects in the environment. This unique representation is partly achieved when features are processed as occurring at the same location. Ideally, the process within the visual system can be seen as doing something like scanning a location and applying a specific tag to the features occurring at that location (maybe by keeping track of that location within object files). For example, all the features occurring at the location i are indexed or tagged with i , and all features occurring at a distinct location m are indexed with m . If the location m and i do not overlap; namely, features in i and m do not occur at the same location, then features are bound into two separate object representations. In real-world perception of cluttered visual scenes, attention serially selects one location after the other, binding the features at each of them. To this extent, the role of attention is to secure identification: it determines when features have a common subject matter and allows for the identification of, and discrimination between, different objects (Clark, 2004a). Object location is, thus, the key element that secures a successful binding of features. This process can be logically characterized as follows:

<red-vertical bar>:
 (at loc_i is R; at loc_j is V; $loc_i = loc_j$ ∴ at loc_i is R and V)
 <green-horizontal bar>:
 (at loc_m is G; at loc_n is H; $loc_m = loc_n$ ∴ at loc_m is G and H)

The logical characterization of visual feature integration has the advantage of outlining the structure of the binding operations. This characterization is an important tool to compare the spatial structure of visual representation with the propositional structure of thought. I argue that the structure of visual representation resembles the structure of constituents of thought.

In fact, the schema above indicates that the representation of an object depends on its constituents being explicitly represented. If not, the derived object representation is only partial. To determine whether vision has a systematic structure of constituents, it is necessary to investigate whether structurally related visual scenes – i.e., scenes that involve different recombinations of objects or features – share the same constituents, and whether visual constituents contribute in the same way, during the binding processes operating on structurally related scenes, to determine the objects of which they are parts. If visual binding mechanisms meet those requirements, then the binding process has a systematic structure of constituents. A systematic recombination of the example visual scene – a green-horizontal bar to the left of a red-vertical bar – requires that at least one of the features belonging to one of the objects in the scene is shifted, so that, as a result, this feature will change its position. Consider a visual scene with a red-horizontal bar to the left of a green-vertical bar. The representation of the example visual scene and the structurally related scene just described can be schematized as follows:

*<green-horizontal bar to the left of a red-vertical bar>:
 (at loc_j is R; at loc_l is V; $loc_j = loc_l$ ∴ at loc_j is R and V)
 (at loc_m is G; at loc_n is H; $loc_m = loc_n$ ∴ at loc_m is G and H)
 **<red-horizontal bar to the left of a green-vertical bar>:
 (at loc_j is R; at loc_k is H; $loc_j = loc_k$ ∴ at loc_j is R and H)
 (at loc_b is G; at loc_c is V; $loc_b = loc_c$ ∴ at loc_b is G and V)

The above configurations show how visual features can be recombined in a systematic fashion by means of combining predicates (features) in a formal language. However, according to Feature Integration Theory, vision does not combine its constituents by means of propositional rules but according to the features' spatial locations. Therefore, it is necessary to provide an argument to explain how visual processes implement the structure just described by means of spatial recombinations.

When two instantiations of the same feature occur at different locations in the world, the feature map coding for that feature will be active. Particularly, it will signal that this specific feature occurs at two distinct locations, corresponding to its locations in the world. In the case of (*) and (**), the same color maps for green and red, and the same orientation maps for horizontal and vertical are active. But the colors are swapped in the two scenes, leading to different object configurations. The difference between the two configurations is encoded in the change of the activated locations in the color maps. The color map signaling green will be active, to simplify, in its "left side" when representing the location of the green feature in scene (*), while it will be active in its "right side" when representing green in scene (**). The converse applies for the feature map coding for red. Thus, whenever two visual scenes are structurally related (as in this case), attentional scanning through the scenes will select object locations, thereby leading to a diverse binding of the features in the structurally related scenes. This results in different object representations in the case of (*) and (**). The binding process is such that primitive constituents are simultaneously tokened with the complex representation. In other words, lacking one of the constituents will result in failure of the binding process.

Thus, feature binding turns out to be more than an associative process that merely links inputs to outputs. In fact, visual binding by spatial attention displays a systematic competency: first, the visual system implements a mereological structure of constituents, rather than processing arbitrarily correlated inputs. Second, the proposed model of visual feature binding displays a systematic structure of constituents. As outlined above, structurally related visual scenes share the same, but differently arranged, primitive features.

Systematicity is a property displayed by both sentential-cognitive representations and spatial representations. This conclusion is in contrast with previous works in philosophy (Clark, 2004b; Fodor, 2008), according to which only representations with a language-like format combine constituents in a way such that a small set of primitive representations can be recombined to form different types of complex representations. In particular, Clark (2004b, p. 571) suggests that sensory states “have something like a subject–predicate structure, though they are not sentential and do not manifest most of the hallmarks of compositionality.” In a classical account, a systematic structure of constituents is a distinctive feature of, and tightly related to, compositionality (Fodor, 1998). The requirement of systematicity is explained in terms of the syntactic structure of constituent recombination in thought, whilst compositionality concerns the content of propositional representations. The main idea is that the content of a thought depends on the content of its constituents and the way they are syntactically combined. The reason Clark argues that visual representations do not have traits that satisfy compositionality is because those representations, arising from the binding of primitive features, provide the basis for the conceptual identification of particulars but do not themselves involve conceptual identification; namely, visual primitive representations do not contribute their content to the content of the final object representation.

I argue, instead, that if a system has a structure of contentful constituents, then this system displays at least one of the hallmarks of compositionality: systematicity. It can also be shown that visual representations satisfy a deflationary notion of compositionality – a weaker form of compositionality than the one mentioned here (Tacca, 2010). A deflationary account only requires that (i) vision has a systematic structure, and that (ii) visual primitive constituents have a specific content. But it remains neutral on which types of semantic properties compose, as required by a classical account of compositionality (Fodor and Lepore, 2001). This is a consequence of the spatial, rather than sentential, character of visual representations. In fact, as Clark notices, visual representations are indeed not sentential. This seems to be the case for both primitive features that are bound at intermediate visual stages and for more complex representations that occur at late visual stages.

The spatial nature of visual representations also makes the systematicity of visual representations different from the systematicity of cognitive representations. The explanation of the systematicity of thought involves two parts (Cummins et al., 2001): (i) it entails that having a thought requires having mental representations that express that thought. This also applies to visual representations, since to represent a visual object, the

primitive representations that code for its characteristics have to be tokened; and (ii) it entails that mental representations have a language-like combinatorial syntax (and semantic). This is not the case for visual representations. Spatial recombinations underlying visual object representation lack the operational repertoire of language-like recombinations. Visual feature binding requires the integration (and spatial grouping) of local, primitive features. To this extent, operations like conjunction and identity are required. But it is not possible to characterize any of the processes involved in binding in terms of other logical operations. No “visual negation” or “visual disjunction” take place. There is no feature integration that is the negation of any of the integrations that occur within the visual system, and, in contrast with feature conjunction, an explicit feature disjunction does not exist in vision: either features are conjoint or they are not combined at all. In sum, vision does not possess the rich propositional structure that higher-cognitive processes seem to have.

The fact that visual representations do not have a propositional nature highlights the difference in combinatorial processes between the visual and perceptual systems but it does not rule out the possibility that systems with different combinatorial structures can implement the same combinatorial requirement, even if in different ways. This is the case for visual representations that, even if they do not allow for propositional recombinations, display systematicity. Thus, the requirement of systematicity can be considered as a general property that does not depend on the type of operations performed on the primitive constituents.

THE CONTENT OF INTERMEDIATE VISUAL REPRESENTATIONS

Another difference between visual perception and cognition concerns the content of their representations. While cognitive representations have conceptual content, the content of perceptual experience is better described as non-conceptual content. Non-conceptual content is often defined in the following way (Bermúdez and Cahen, 2011): a mental state has non-conceptual content if and only if the subject of that state does not need to possess the relevant concepts required to specify its content.

How to define then the non-conceptual content of perceptual states? Heck (2007) argues that what kind of content perceptual and cognitive states have is a question about what kinds of representations those states involve. Heck’s analysis starts from the premise that the conceptual content of beliefs is structured in a way that fulfills the requirement of the Generality Constraint. The debate over non-conceptual content then turns out to be about whether the cognitive abilities one exercises when one thinks that tomatoes are red are also exercised when one veridically perceive a ripe tomato, and whether it would be impossible for one to perceive the tomato as one does were one not able to think as one can. Thus, the question of what kind of content one should take perceptual experience to have has to be answered by investigating the structural characteristics of perceptual representations. The content of perception will be conceptual only if the Generality Constraint is satisfied (Heck, 2007). But, according to Heck, this is not the case,

since visual representations, as described by pictorial theories, have a spatial structure that violates even the weaker requirement of systematicity. Satisfying systematicity is a necessary condition on satisfying the Generality Constraint. Therefore, since visual representations do not display systematicity, their content is non-conceptual.

The analysis proposed in this paper of how visual representations spatially combine leads, instead, to a different conclusion: the appeal to the spatial structure of vision seems to count in favor of the conceptualist thesis, rather than providing a strong argument for the existence of representations with a non-conceptual content. This is because visual representations satisfy the requirement of systematicity – i.e., a necessary condition to satisfy the Generality Constraint. Systematicity is a weak-syntactic reading of the Generality Constraint that states that there is a certain kind of pattern in our cognitive capacities. In this form, the requirement of systematicity describes representational composites as depending on syntactic recombinations involving the same constituents. Recombinations of cognitive representations entail that a person has conceptual abilities (McLaughlin, 2009). In the case of visual perception, systematic recombinations of primitive features involve the ability of a subject to identify particular features. This ability might correspond to an *early* type of conceptual ability, since visual representations, like cognitive representations, are constituted by primitive constituents that make the same contribution in structurally related representations. Particularly, the representation of features within feature maps is such that whenever a feature is tokened in the feature map (e.g., “red”), this feature will contribute in the same way to the final object representations in which the color red is involved (e.g., a red-vertical bar, a red-horizontal bar). While the contribution of the feature representation is the same in different object representations, those representations will differ from each other as a function of the spatial configuration of their features, since, for different object representations, feature locations are different. This is similar to what occurs in propositional representations, for which, although the same constituent (e.g., the concept RED) contributes in the same way to thoughts regarding red things, the final complex representations depend on the syntactic configurations of the primitive constituents.

However, unlike propositional representations, the possession of systematic perceptual skills is not sufficient to satisfy the Generality Constraint in its strong form, and, thus, not enough to establish both necessary and sufficient conditions for the conceptuality of perceptual representations. The idea behind the Generality Constraint is that conceptual representations involve not only a systematic recombination of primitive constituents but also an abstract grasp on the way things are. Thought representations, and propositional representations in general, are not constrained to any mode of access (Peacocke, 2001). We can, in principle, entertain an indefinite number of thoughts. This is based on the idea that human thoughts have an unbound competence that is not limited by our performance (Fodor and Pylyshyn, 1988; Tacca, 2010). Instead, our perceptual representation of the world is bound to the limit of the perceptual system in use. We cannot perceive an indefinite number of visual scenes, since what we can perceive depends on the physical constitution of our visual system. There is no such thing as an abstract visual competence.

Nevertheless, it can be argued that failure to satisfy the Generality Constraint in its fullest version – that is, by showing both systematic combinability and abstract competence – does not exclude intermediate visual representations from being a specific type of conceptual representations. Perceptual representations might count as an early type of conceptual representations that will become more abstract only with full possession of conceptual resources. These early types of conceptual representations display systematic recombination – a necessary even if not sufficient condition for a person to possess conceptual abilities. Moreover, the definition of visual representations as early types of conceptual representations will also provide a link between human and animal cognition. Some of the criteria analyzed here, particularly systematicity, have been reported as basic criteria for showing concept possession in animals, too (Newen and Bartels, 2007). Thus, the distinction of the content of perception and cognition based on satisfaction of systematicity does not show that the content of conscious perceptual experience is nonconceptual. At best, one can argue that satisfaction of the requirement of systematicity shows that intermediate stage visual representations, the ones involved in the binding process, might be an early type of conceptual representations. The abstract grasp on ways of representing the world, required by the full satisfaction of the Generality Constraint, is then a criterion to distinguish fully conceptual–cognitive representations from early types of conceptual–perceptual representations; rather than to distinguish conceptual from non-conceptual representations. However, while visual representations at intermediate stages have properties that characterize their content as conceptual, it is still possible that visual representations at early visual stages (e.g., feature segregation, boundary representation) have non-conceptual content. At this stage, there is hardly any influence from cognitive processes, and recombination of primitive constituents that satisfy the requirement of systematicity does not seem to occur. Thus, it might be that the transition between representations with nonconceptual and conceptual content occurs already between early and intermediate visual stages.

CONCLUDING REMARKS

To claim that perception and cognition are tightly related makes sense only if one can explain how those systems, which are individuated in different brain areas and process different types of information, communicate. In this paper, I argue that visual representations share a structural property with cognitive representations; namely, that spatial recombination of visual representations into an object representation displays systematicity. This conclusion contrasts the traditional view in philosophy, according to which only sentential-cognitive representations implement a systematic structure of constituents, and it is in line with findings in physiology and psychology of how the visual system creates object representations.

The fact that visual perceptual representation, even if not sentential, displays systematicity poses a further problem for philosophical theories that see systematicity as a hallmark of representations with conceptual content. I argue that if one takes the satisfaction of this requirement as a necessary condition for having conceptual content then the content of visual representations

amounts to an early type of conceptual content that does not allow for the same kind of abstraction that is typical of human cognitive abilities. This type of early conceptual and perceptual content might be a characteristic that humans have in common with animals.

Moreover, showing that visual representations display systematicity makes it easier to see how visual perception and cognition might relate and share representational information. In fact, one of the problems of claiming that visual perception and cognition have different structure and content is that it becomes unclear how they can share information. It might be that implementing a systematic structure is a basic way of recombination that is shared by different brain areas. This might be a general code of assembling information that makes more efficient its processing in different modalities.

REFERENCES

- Barrett, H. C. (2005). Enzymatic computation and cognitive modularity. *Mind Lang.* 259–287.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 577–660.
- Bermúdez, J. L. (2007). What is at stake in the debate on nonconceptual content? *Philos. Perspect.* 21, 55–72.
- Bermúdez, J. L., and Cahen, A. (2011). “Nonconceptual mental content,” in *Stanford Encyclopedia of Philosophy*, Summer 2011 Edn, ed. E. N. Zalta. Available at: <http://plato.stanford.edu/entries/content-nonconceptual/>
- Blaser, E., Pylyshyn, Z. W., and Holcombe, A. O. (2000). Tracking an object through feature space. *Nature* 408, 196–199.
- Block, N. (2005). Two neural correlates of consciousness. *Trends Cogn. Sci. (Regul. Ed.)* 9, 46–52.
- Brewer, W. F., and Lambert, B. L. (2001). The theory-ladenness of observation and the theory-ladenness of the rest of the scientific process. *Philos. Sci.* 3, 176–186.
- Camp, E. (2009). Putting thoughts to work: concepts, systematicity, and stimulus-independence. *Philos. Phenomenol. Res.* 78, 275–311.
- Clark, A. (2004a). Feature-placing and proto-objects. *Philos. Psychol.* 17, 443–469.
- Clark, A. (2004b). Sensing, objects, and awareness: reply to commentators. *Philos. Psychol.* 17, 553–579.
- Cummins, R., Blackmon, J., Byrd, D., Poirier, P., Roth, M., and Schwarz, G. (2001). Systematicity and the cognition of structured domains. *J. Philos.* 98, 167–185.
- Engel, A. K., Kreiter, A. K., Konig, P., and Singer, W. (1991). Synchronization of oscillatory neuronal responses between striate and extrastriate visual cortical areas of the cat. *Proc. Natl. Acad. Sci. U.S.A.* 88, 6048–6052.
- Evans, G. (1982). *The Varieties of Reference*. Oxford: Oxford University Press.
- Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47.
- Fodor, J. A. (1975). *The Language of Thought*. Cambridge, MA: Harvard University Press.
- Fodor, J. A. (1998). *Concepts: Where Cognitive Science Went Wrong*. Oxford: Clarendon Press.
- Fodor, J. A. (2007). “The revenge of the given,” in *Contemporary Debates in Philosophy of Mind*, eds B. P. McLaughlin and J. Cohen (New York: Blackwell Publishing), 105–116.
- Fodor, J. A. (2008). *LOT 2: The Language of Thought Revisited*. Oxford: Oxford University Press.
- Fodor, J. A., and Lepore, E. (2001). Why compositionality won’t go away: reflections on Horwich’s ‘deflationary’ theory. *Ratio* 14, 350–368.
- Fodor, J. A., and McLaughlin, B. P. (1990). Connectionism and the problem of systematicity: why Smolensky’s solution doesn’t work. *Cognition* 35, 183–204.
- Fodor, J. A., and Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: a critical analysis. *Cognition* 28, 3–71.
- Fries, P., Reynolds, J. H., Rorie, A. E., and Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science* 291, 1560–1563.
- Goldstone, R., and Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition* 65, 231–262.
- Heck, R. G. (2007). “Are there different kinds of content?” in *Contemporary Debates in Philosophy of Mind*, eds B. P. McLaughlin and J. Cohen (New York: Blackwell Publishing), 117–138.
- Jackson, F. (1977). *Perception: A Representative Theory*. Cambridge: Cambridge University Press.
- Kahneman, D., Treisman, A., and Gibbs, B. J. (1992). The reviewing of object files: object-specific integration of information. *Cogn. Psychol.* 24, 175–219.
- Koch, C., and Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Hum. Neurobiol.* 4, 219–227.
- Lamme, V. A. F. (2003). Why visual attention and awareness are different. *Trends Cogn. Sci. (Regul. Ed.)* 7, 12–18.
- Livingstone, M., and Hubel, D. (1988). Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science* 240, 740–749.
- Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: Freeman.
- McLaughlin, B. P. (2009). Systematicity redux. *Synthese* 170, 251–274.
- Newen, A., and Bartels, A. (2007). Animal minds and the possession of concepts. *Philos. Psychol.* 20, 283–308.
- Peacocke, C. (2001). Does perception have a nonconceptual content? *J. Philos.* 98, 239–264.
- Pecher, D., Zeelenberg, R., and Barsalou, L. W. (2004). Sensorimotor simulations underlie conceptual representations: modality-specific effect of prior act. *Psychon. Bull. Rev.* 11, 164–167.
- Prinz, J. (2002). *Furnishing the Mind: Concepts and Their Perceptual Basis*. Cambridge, MA: MIT Press.
- Raftopoulos, A., and Mueller, V. C. (2006). The phenomenal content of experience. *Mind Lang.* 21, 187–219.
- Raftopoulos, A. (this issue). Late vision: processes and epistemic status. *Front. Psychol.*
- Reverberi, C., Gorgen, K., and Haynes, J. D. (2011). Compositionality of rule representations in human prefrontal cortex. *Cereb. Cortex*. doi: 10.1093/cercor/bhr200. [Epub ahead of print].
- Robertson, L. C. (2003). Binding, spatial attention, and perceptual awareness. *Nat. Rev. Neurosci.* 4, 93–102.
- Roskies, A. L. (1999). The binding problem. *Neuron* 24, 111–125.
- Scholl, B. J. (2009). “What have we learned about attention from multiple-object tracking (and vice versa)?” in *Computation, Cognition, and Pylyshyn*, eds D. Dedrick and L. Trick (Cambridge, MA: MIT Press), 49–78.
- Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist networks. *Artif. Intell.* 46, 59–216.
- Tacca, M. C. (2010). *Seeing Objects: The Structure of Visual Representations*. Paderborn: Mentis.
- Toribio, J. (2008). State versus content: the unfair trial of perceptual nonconceptualism. *Erkenntnis* 69, 351–361.
- Treisman, A. M. (1993). “The perception of features and objects,” in *Attention: Selection, Awareness, and Control*, eds A. Baddeley and L. Weiskrantz (Oxford: Clarendon Press), 1–36.
- Treisman, A. M. (1996). The binding problem. *Curr. Opin. Neurobiol.* 6, 171–178.

ACKNOWLEDGMENTS

This work was funded by the German Research Foundation (DFG-FOR600). I thank Tobias H. Donner and two reviewers for comments.

- Treisman, A. M., and Gelade, G. (1980). A feature-integration theory of attention. *Cogn. Psychol.* 12, 97–136.
- Treisman, A. M., and Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cogn. Psychol.* 14, 107–141.
- Wolfe, J. M. (1998). “Visual search,” in *Attention*, ed. H. Pashler (London: University College London Press), 13–73.
- Wolfe, J. M., and Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nat. Rev. Neurosci.* 5, 495–501.
- Zeki, S. (1978). Functional specialization in the visual cortex of the rhesus monkey. *Nature* 274, 423–428.
- Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received: 06 September 2011; paper pending published: 23 September 2011; accepted: 14 November 2011; published online: 30 November 2011.
- Citation: Tacca MC (2011) Commonalities between perception and cognition. *Front. Psychology* 2:358. doi: 10.3389/fpsyg.2011.00358
- This article was submitted to *Frontiers in Perception Science*, a specialty of *Frontiers in Psychology*.
- Copyright © 2011 Tacca. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits use, distribution, and reproduction in other forums, provided the original authors and source are credited.